A Brief Introduction Into Activation-Based Conditional Inference

 $\begin{array}{c} {\rm Marco\ Wilhelm^{1[0000-0003-0266-2334]},\ Diana\ Howey^{1[0000-0002-7203-4862]},} \\ {\rm Gabriele\ Kern-Isberner^{1[0000-0001-8689-5391]},\ Kai} \end{array} ,$

Sauerwald²[0000-0002-1551-7016], and Christoph Beierle²[0000-0002-0736-8516]

¹ Dept. of Computer Science, TU Dortmund University, Dortmund, Germany {marco.wilhelm, diana.howey, gabriele.kern-isberner}@cs.tu-dortmund.de

² Dept. of Computer Science, FernUniversität in Hagen, Hagen, Germany

{kai.sauerwald, christoph.beierle}@fernuni-hagen.de

Abstract. Activation-based conditional inference integrates several aspects of human reasoning into formal conditional reasoning, such as focusing, forgetting, and remembering, by combining conditional reasoning and the cognitive architecture ACT-R. The idea is to select a reasonable subset of a conditional belief base before drawing inferences. The selection is based on an activation function which assigns to the conditionals in the belief base a degree of activation based on the conditional's relevance for the current query and its usage history.

1 Introduction

Activation-based conditional inference combines ACT-R [2, 1] and conditional reasoning. ACT-R (Adaptive Control of Thought-Rational) is a well-founded cognitive architecture developed to formalize human reasoning in which a selection of cognitive entities (chunks as declarative memory and production rules as procedural memory) is performed before these entities are used to solve a reasoning task. From a cognitive point of view, there are basically two processes which affect the selection: The long-term process of forgetting and remembering and the short-term process of activating certain beliefs depending on the current context. In this paper, we adapt the concept of (de)activation of cognitive entities from ACT-R and combine it with the task of drawing conditional inferences. More precisely, we define an activation function for conditionals of the form (B|A) with the meaning "if A holds, then usually B holds, too." The conditionals with the highest activation are selected for the inference task. Therewith, we generalize the concept of focused inference [6] and give it a profound cognitive meaning, and we also equip ACT-R with a modern, high quality inference formalism.

2 Logical Foundations

We consider a propositional language \mathcal{L} over a finite set of atoms Σ which we extend to the conditional language $(\mathcal{L}|\mathcal{L}) = \{(B|A) \mid A, B \in \mathcal{L}\}$ where conditionals

Copyright © 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

A Brief Introduction Into Activation-Based Conditional Inference

 $(B|A) \in (\mathcal{L}|\mathcal{L})$ have the intuitive meaning "If A holds, then usually B holds, too." A formal semantics of conditionals is given by ranking functions over possible worlds [5]. Here, possible worlds are the propositional interpretations $I \in \mathcal{I}$ represented as complete conjunctions of literals (atoms or their negations). The set of all possible worlds is denoted by Ω . A ranking function $\kappa : \Omega \to \mathbb{N}_0^{\infty}$ with $\kappa^{-1}(0) \neq \emptyset$ maps possible worlds to a degree of plausibility. Lower ranks indicate higher plausibility so that $\kappa^{-1}(0)$ is the set of the most plausible worlds. Ranking functions are extended to formulas by $\kappa(A) = \min\{\kappa(\omega) \mid \omega \models A\}$. κ accepts a conditional (B|A) iff $\kappa(AB) < \kappa(A\overline{B})$ or $\kappa(A) = \infty$ and is a (ranking) model of a belief base Δ (a finite set of conditionals) iff κ accepts all conditionals in Δ .

An inference operator [3] is a mapping \mathfrak{I} which assigns to each belief base Δ an inference relation $\succ^{\mathfrak{I}}_{\Delta} \subseteq \mathcal{L} \times \mathcal{L}$ such that

$-$ if $(B A) \in \Delta$, then $A \models^{\mathfrak{I}}_{\Delta} B$,	(Direct Inference)
- if $\Delta = \emptyset$, then $A \models^{\mathfrak{I}}_{\Delta} B$ only if $A \models B$.	(Trivial Vacuity)

 $\mathfrak{I}_{\Delta} = \{(B|A) \mid A \models_{\Delta}^{\mathfrak{I}} B\}$ denotes the set of *inductive inferences* from Δ wrt. \mathfrak{I} . Inference operators yield a three-valued *inference response* to a *query* (B|A):

$$\llbracket (B|A) \rrbracket_{\Delta}^{\mathfrak{I}} = \begin{cases} yes & \text{iff} \quad (B|A) \in \mathfrak{I}_{\Delta} \\ no & \text{iff} \quad (\overline{B}|A) \in \mathfrak{I}_{\Delta} \\ unknown & \text{otherwise} \end{cases}$$

Drawing inferences from the whole belief base Δ can be computationally expensive and does not fit to human reasoning. Thus, focused inference [6] defines a (query-dependent) subset $\phi(\Delta) \subseteq \Delta$ as a focus and draws inferences wrt. $\phi(\Delta)$ instead of Δ : A conditional (B|A) follows from Δ wrt. the inference operator \Im in the focus $\phi(\Delta)$ iff $(B|A) \in \Im_{\phi(\Delta)}$. Finding an appropriate focus is challenging but, apart from the computational benefits of small foci, appropriate foci may unveil the part of Δ which is relevant for answering the query.

3 ACT-R Architecture

ACT-R [2, 1] is a cognitive architecture which formalizes human reasoning and distinguishes between *declarative* and *procedural memory*. In the declarative memory, categorical knowledge about individuals or objects is stored in form of *chunks* while the procedural memory consists of *production rules* and describes how chunks are processed. Reasoning in ACT-R starts with an initial *priming* of chunks. The chunk with the highest activation is processed by production rules in order to compute a solution to the reasoning task. If this fails, the activation passes into an iterative process. The retrieval of chunks basically depends on an *activation function* which is calculated for each specific request anew and is given by the *base-level activation* $\mathcal{B}(\mathbf{c}_i)$ and the *spreading activation* $\mathcal{S}(\mathbf{c}_i)$, which is a sum of *degrees of associations between chunks* $\mathcal{S}(\mathbf{c}_i, \mathbf{c}_j)$ weighted by $\mathcal{W}(\mathbf{c}_j)$:

$$\mathcal{A}(\mathfrak{c}_i) = \mathcal{B}(\mathfrak{c}_i) + \mathcal{S}(\mathfrak{c}_i) = \mathcal{B}(\mathfrak{c}_i) + \sum_j \mathcal{W}(\mathfrak{c}_j) \cdot \mathcal{S}(\mathfrak{c}_i, \mathfrak{c}_j).$$
(1)

The base-level activation of a chunk $\mathcal{B}(\mathbf{c}_i)$ reflects the entrenchment of \mathbf{c}_i in the reasoner's memory and depends on the recency and frequency of its use. Typically, $\mathcal{B}(\mathbf{c}_i)$ is decreased over time (fading out) and is increased when the chunk is active. The spreading activation of a chunk $\mathcal{S}(\mathbf{c}_i)$ formalizes the impact of the priming. While the degree of association $\mathcal{S}(\mathbf{c}_i, \mathbf{c}_j)$ reflects how strongly related the two chunks \mathbf{c}_i and \mathbf{c}_j are in principal (i.e., it reflects whether \mathbf{c}_i and \mathbf{c}_j deal with the same issue or not), the weighting factor $\mathcal{W}(\mathbf{c}_i)$ indicates whether this connection is triggered by the actual priming.

4 Activation-Based Conditional Inference

The production-system-based logical basis of ACT-R does not hold the pace with modern KRR formalisms. Therefore, we propose a cognitively inspired model of conditional reasoning by interpreting the concepts of ACT-R in terms of logic, conditionals, and inference. In particular, we replace chunks by conditionals in a belief base Δ and derive a focus $\phi(\Delta)$ based on the activation of the conditionals in order to draw focused inferences. Atoms in \mathcal{L} play the role of cognitive units, and the production rules are replaced by an inference operator \Im . From the conditional logical perspective, the main value of this approach is the cognitive justification of the focus and the option to integrate further cognitive concepts such as forgetting and remembering. More formally, we calculate an activation value $\mathcal{A}(\mathfrak{r}) > 0$ for all $\mathfrak{r} \in \Delta$. If $\mathcal{A}(\mathfrak{r})$ is not less than a certain threshold $\theta \geq 0$, then the conditional \mathfrak{r} is within the focus $\phi(\Delta)$, i.e.,

$$\phi(\Delta) = \phi(\Delta, \mathcal{A}, \theta) = \{ \mathfrak{r} \in \Delta \mid \mathcal{A}(\mathfrak{r}) \ge \theta \}.$$

Note that $\phi(\Delta)$ implicitly depends on a query $\mathbf{q} = (B|A)$, too, since queries will serve as the priming and \mathcal{A} depends on that priming. Eventually, we say that the query \mathbf{q} can be *inferred from* Δ *wrt.* \Im , \mathcal{A} , and θ iff $\mathbf{q} \in \Im_{\phi(\Delta,\mathcal{A},\theta)}$. When answering the query fails, i.e., if $[\![\mathbf{q}]\!]_{\phi(\Delta,\mathcal{A},\theta)}^{\Im} = unknown$, the inference process can be iterated by lowering the threshold θ . In the limit, when $\theta = 0$, one has $\phi(\Delta,\mathcal{A},0) = \Delta$, thus $[\![\mathbf{q}]\!]_{\phi(\Delta,\mathcal{A},0)}^{\Im} = [\![\mathbf{q}]\!]_{\Delta}^{\Im}$.

In the ACT-R framework, the functionality of the activation function (1) is extensively discussed but its single components are not formalized mathematically. Here, we give a concrete instantiation of (1) in the conditional setting which can be seen as a blue print for further investigations and empirical analyses. Let Δ be a belief base, $\mathfrak{r}_i \in \Delta$, and \mathfrak{q} a further conditional (the query resp. priming), then the *activation* of \mathfrak{r}_i wrt. Δ and \mathfrak{q} is

$$\mathcal{A}_{\mathfrak{q}}^{\Delta}(\mathfrak{r}_{i}) = \underbrace{\mathcal{B}^{\Delta}(\mathfrak{r}_{i})}_{\text{base-level activation}} + \underbrace{\sum_{\mathfrak{r}_{j} \in \Delta} \mathcal{W}_{\mathfrak{q}}^{\Delta}(\mathfrak{r}_{j}) \cdot \mathcal{S}(\mathfrak{r}_{i}, \mathfrak{r}_{j})}_{\text{spreading activation } \mathcal{S}_{\mathfrak{q}}^{\Delta}(\mathfrak{r}_{i})}.$$
 (2)

In (2), the base-level activation $\mathcal{B}^{\Delta}(\mathfrak{r})$ reflects the entrenchment of \mathfrak{r} in the reasoner's memory. Since epistemic entrenchment and ranking semantics are dual

ratings, the *normality* of a conditional is a good estimator and we define

$$\mathcal{B}^{\Delta}(\mathfrak{r}) = rac{1}{1 + Z^{\Delta}(\mathfrak{r})}, \qquad \mathfrak{r} \in \Delta$$

where $Z^{\Delta}(\mathfrak{r})$ is the *Z*-rank of \mathfrak{r} from System Z which is a valuable measure of normality according to [4].

The degree of association $\mathcal{S}(\mathfrak{r}_i,\mathfrak{r}_j)$ is a measure of connectedness between the conditionals in Δ and is defined by

$$\mathcal{S}(\mathfrak{r}_i,\mathfrak{r}_j) = \frac{|\Sigma(\mathfrak{r}_i) \cap \Sigma(\mathfrak{r}_j)|}{|\Sigma(\mathfrak{r}_i) \cup \Sigma(\mathfrak{r}_j)|}, \qquad \mathfrak{r}_i, \mathfrak{r}_j \in \Delta,$$

where $\Sigma(\mathfrak{r})$ is the set of atoms mentioned in \mathfrak{r} . That is, $\mathcal{S}(\mathfrak{r}_i, \mathfrak{r}_j)$ is the number of shared atoms relative to all atoms in \mathfrak{r}_i or \mathfrak{r}_j . This syntactically-driven definition of $\mathcal{S}(\mathfrak{r}_i, \mathfrak{r}_j)$ is motivated by and extends the *principle of relevance* from nonmonotonic reasoning which states that if the belief base Δ splits into two sub-belief bases Δ_1 and Δ_2 with $\Sigma(\Delta_1) \cap \Sigma(\Delta_2) = \emptyset$ and the query is defined over one of the signatures $\Sigma(\Delta_i)$, say $\Sigma(\Delta_1)$, only, then only the conditionals in Δ_1 should be relevant for answering this query (cf. [3]).

The weighting factor $\mathcal{W}_{q}^{\Delta}(\mathfrak{r})$ indicates how much the priming \mathfrak{q} triggers the conditional \mathfrak{r} . We formalize the influence of the priming according to the *spread*ing activation theory [1] by a labeling of the vertices in an undirected graph $\mathcal{N}(\Delta)$ with vertices $\mathcal{V} = \Sigma$ and edges

$$\mathcal{E} = \{\{\mathfrak{a}, \mathfrak{b}\} \mid \exists \mathfrak{r} \in \Delta : \{\mathfrak{a}, \mathfrak{b}\} \subseteq \Sigma(\mathfrak{r})\}.$$

The labels are the *triggering values* $\tau_{\mathfrak{q}}^{\Delta}(\mathfrak{a}) \in [0, 1]$ for $\mathfrak{a} \in \Sigma$ which indicate how much \mathfrak{a} is triggered by \mathfrak{q} . Once the triggering values are determined, we follow the idea that a conditional \mathfrak{r} is triggered not more than the atoms in $\Sigma(\mathfrak{r})$ and define the respective weighting factor by

$$\mathcal{W}^{\Delta}_{\mathfrak{a}}(\mathfrak{r}) = \min\{\tau^{\Delta}_{\mathfrak{a}}(\mathfrak{a}) \mid \mathfrak{a} \in \Sigma(\mathfrak{r})\}.$$

The actual labeling of the vertices in $\mathcal{N}(\Delta)$ is an iterative process and starts with labeling the atoms $\mathfrak{a} \in \Sigma$ which are mentioned in the query \mathfrak{q} , i.e. $\mathfrak{a} \in \Sigma(\mathfrak{q})$, with $\tau_{\mathfrak{q}}^{\Delta}(\mathfrak{a}) = 1$. In the subsequent steps, the neighboring atoms are labeled and so on. The remaining atoms \mathfrak{a}' which are not reachable from the initially labeled atoms in $\Sigma(\mathfrak{q})$ are labeled with $\tau_{\mathfrak{q}}^{\Delta}(\mathfrak{a}') = 0$. The labels of the atoms \mathfrak{a}'' in between are the sum of the labels of the already labeled neighbors weighted by the sum of all labels so far plus 1, i.e.,

$$\tau_{\mathfrak{q}}^{\Delta}(\mathfrak{a}'') = \frac{\sum_{\mathfrak{b}\in\mathcal{L}:\,\{\mathfrak{a}'',\mathfrak{b}\}\in\mathcal{E}} \tau_{\mathfrak{q}}^{\Delta}(\mathfrak{b})}{1 + \sum_{\mathfrak{b}\in\mathcal{L}} \tau_{\mathfrak{a}}^{\Delta}(\mathfrak{b})}$$

where \mathcal{L} is the set of the already labeled atoms. This guarantees that these labels are between 0 and 1 and decrease for increasing iteration steps. Therewith, the triggering value of an atom depends on both the triggering values of the associated (earlier triggered) atoms and their quantity.

5 Forgetting and Remembering

In *ACT-R* the base-level activation of a chunk is not constant but decreases over time and increases when the chunk is retrieved. This integrates the concepts of forgetting and remembering into *ACT-R*. In order to capture this dynamic view on the base-level activation, we propose to extend the base-level activation such that $\mathcal{B}^{\Delta}(\mathbf{r})$ is multiplied with a *forgetting factor* after each inference request. For a fixed $\delta \in [0, 1)$, the focus $\phi(\Delta) = \phi(\Delta, \mathcal{A}, \theta)$ and $\mathbf{r} \in \Delta$, the forgetting factor $\Phi_{\delta,\phi(\Delta)}(\mathbf{r})$ is given by $1 + \delta$ if $\mathbf{r} \in \phi(\Delta)$ and otherwise given by $1 - \delta$. By doing so, the base-level activation of a conditional is decreased when the conditional is not selected for answering the query, and it is increased otherwise. When applying this update of the base-level activation for every inference request, the usage history of the conditionals is implemented into $\mathcal{B}^{\Delta}(\mathbf{r})$.

6 Conclusions and Future Work

We applied conditional reasoning to ACT-R [2, 1] and developed a prototypical model for activation-based conditional inference. For this, we reformulated the activation function from ACT-R for conditionals and selected the conditionals with the highest degree of activation for drawing inference. With our approach it is possible to implement several aspects of human reasoning into modern expert systems such as focusing, forgetting, and remembering. The main challenge for future work is to find for a given query $\mathbf{q} = (B|A)$ a proper subset Δ' of a belief base Δ such that \mathbf{q} is answered the same wrt. to Δ' and Δ , i.e., $\llbracket \mathbf{q} \rrbracket_{\Delta'}^2 = \llbracket \mathbf{q} \rrbracket_{\Delta}^2$.

Acknowledgments. This work is supported by DFG Grant KE 1413/10-1 awarded to Gabriele Kern-Isberner and DFG Grant BE 1700/9-1 awarded to Christoph Beierle as part of the priority program "Intentional Forgetting in Organizations" (SPP 1921).

References

- 1. Anderson, J.R.: How can the human mind occur in the physical universe? Oxford University Press (2007)
- Anderson, J.R., Lebiere, C.: The atomic components of thought. Psychology Press (1998)
- Kern-Isberner, G., Beierle, C., Brewka, G.: Syntax splitting = relevance + independence: New postulates for nonmonotonic reasoning from conditional belief bases. In: Proceedings of the 17th International Conference on Principles of Knowledge Representation and Reasoning. pp. 560–571 (2020)
- Pearl, J.: System Z: A natural ordering of defaults with tractable applications to nonmonotonic reasoning. In: Proceedings of the 3rd Conference on Theoretical Aspects of Reasoning about Knowledge. pp. 121–135. Morgan Kaufmann (1990)
- 5. Spohn, W.: The Laws of Belief: Ranking Theory and Its Philosophical Applications. Oxford University Press (2012)
- Wilhelm, M., Kern-Isberner, G.: Focused inference and System P. In: Thirty-Fifth AAAI Conference on Artificial Intelligence. pp. 6522–6529. AAAI Press (2021)